



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
-----------------	-------------	----------------------	---------------------	------------------

09/844,120

04/27/2001

Sara H. Basson

YOR920010063US1

1052

7590

09/02/2005

Ryan, Mason & Lewis, LLP  
90 Forest Avenue  
Locust Valley, NY 11560

EXAMINER

VO, HUYEN X

ART UNIT

PAPER NUMBER

2655

DATE MAILED: 09/02/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

## Office Action Summary

**Application No.**

09/844,120

**Applicant(s)**

BASSON ET AL.

**Examiner**

Huyen X. Vo

**Art Unit**

2655

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --  
**Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

### Status

- 1) ☒ Responsive to communication(s) filed on 17 June 2005.  
2a) ☐ This action is **FINAL**. 2b) ☒ This action is non-final.  
3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

### Disposition of Claims

- 4) ☒ Claim(s) 1-24 is/are pending in the application.  
4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.  
5) ☐ Claim(s) \_\_\_\_\_ is/are allowed.  
6) ☒ Claim(s) 1-24 is/are rejected.  
7) ☐ Claim(s) \_\_\_\_\_ is/are objected to.  
8) ☐ Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

### Application Papers

- 9) ☐ The specification is objected to by the Examiner.  
10) ☒ The drawing(s) filed on 27 April 2001 is/are: a) ☒ accepted or b) ☐ objected to by the Examiner.  
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).  
11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

### Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).  
a) ☐ All b) ☐ Some \* c) ☐ None of:  
1. ☐ Certified copies of the priority documents have been received.  
2. ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.  
3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

\* See the attached detailed Office action for a list of the certified copies not received.

### Attachment(s)

- 1) ☒ Notice of References Cited (PTO-892)  
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)  
3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)  
Paper No(s)/Mail Date \_\_\_\_\_.  
4) ☐ Interview Summary (PTO-413)  
Paper No(s)/Mail Date \_\_\_\_\_.  
5) ☐ Notice of Informal Patent Application (PTO-152)  
6) ☐ Other: \_\_\_\_\_.

## **DETAILED ACTION**

### ***Claim Rejections - 35 USC § 112***

1. The following is a quotation of the second paragraph of 35 U.S.C. 112:

The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

2. Claims 1-24 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

3. The term "substantially" in claims 1, 10, 13, 18, and 24 is a relative term which renders the claim indefinite. The term "substantially" is not defined by the claim, the specification does not provide a standard for ascertaining the requisite degree, and one of ordinary skill in the art would not be reasonably apprised of the scope of the invention. Examiner does not include the term "substantially" in the interpretation of the claim language. Appropriate correction is required.

### ***Claim Rejections - 35 USC § 102***

4. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless – (e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

5. Claims 1-3, 10-15, 17-21, and 23-24 are rejected under 35 U.S.C. 102(e) as being anticipated by Braida et al. (US 6317716).

6. Regarding claim 1, Braida et al. disclose an apparatus for presenting images representative of one or more words in an utterance with corresponding decoded speech, the apparatus comprising: a visual detector, the visual detector capturing images of body movements concurrently from the one or more words in the utterance (*camera 8 in figure 2 record images of body movement concurrently with input speech recorded by the microphone 9*); a visual feature extractor coupled to the visual detector (*the main processor 34 in figure 2*), the visual feature extractor receiving time information from an automatic speech recognition (ASR) system and operatively processing the captured images into one or more image segments based on the time information relating to one or more words, decoded by the ASR system, in the utterance, each image segment comprising a plurality of successive images in time corresponding to a decoded word in the utterance (*col. 6, lines 58-67, timestamp is generated by the phone recognition. Col. 10, lines 30-47, received images of body movement is processed to include decoded words (cue images) according to the timestamp*); and an image player operatively coupled to the visual feature extractor, the image player receiving and presenting decoded word with each image segment generated therefrom (*Display 18 in figure 2*).

7. Regarding claim 10, Braida et al. disclose an apparatus for presenting images representative of one or more words in an utterance with corresponding decoded speech, the apparatus comprising: an automatic speech recognition (ASR) engine for converting the utterance into the one or more decoded words, the ASR engine generating time information associated with each of the decoded words (*elements 42-50 in figure 3 and/or referring to col. 6, lines 53-67 and col. 10, lines 30-47, decoded words being cue images*); a visual detector, the visual detector capturing images of body movements concurrently from one or more words in the utterance (*camera 8 in figure 2 record images of body movement concurrently with input speech recorded by the microphone 9*); a visual feature extractor coupled to the visual detector (*the main processor 34 in figure 2*), the visual feature extractor receiving the time information from the ASR engine and operatively processing the captured images into one or more image segments based on the time information relating to the decoded words, each image segment comprising a plurality of successive images in time corresponding to a decoded word in the utterance (*col. 6, lines 58-67, timestamp is generated by the phone recognition. Col. 10, lines 30-47, received images of body movement is processed to include decoded words (cue images) according to the timestamp*); and an image player operatively coupled to the visual feature extractor, the image player receiving and presenting decoded word with each image segment generated therefrom (*Display 18 in figure 2*).

8. Regarding claims 13 and 18, Braida et al. disclose a method for presenting images representative of one or more words in an utterance with corresponding decoded speech, the method comprising the steps of: capturing a plurality of images representing body movements concurrently from the one or more words in the utterance (*camera 8 in figure 2 record images of body movement concurrently with input speech recorded by the microphone 9*); associating each of the captured images generated from the one or more words in the utterance with time information relating to an occurrence of the image (*col. 6, lines 58-67, timestamp is generated by the phone recognition. Col. 10, lines 30-47, received images of body movement is processed to include decoded words (cue images) according to the timestamp*); receiving, from an automatic speech recognition (ASR) system, data including a start time and an end time of a word decoded by the ASR system (*col. 6, lines 58-67, timestamp is generated by the phone recognition. Col. 10, lines 30-47, received images of body movement is processed to include decoded words (cue images) according to the timestamp*); aligning the plurality of images into one or more image segments according to the start and stop times received from the ASR system, wherein each image segment corresponds to a decoded word in the utterance (*col. 6, lines 58-67, timestamp is generated by the phone recognition. Col. 10, lines 30-47, received images of body movement is processed to include decoded words (cue images) according to the timestamp*); and presenting an image segment with a corresponding decoded word (*Display 18 in figure 2*).

Art Unit: 2655

9. Regarding claim 24, Braida et al. disclose a method for presenting images representative of one or more words in an utterance with corresponding decoded speech, the method comprising the steps of: providing an automatic speech recognition (ASR) engine (*elements 42-50 in figure 3*); decoding, in the ASR engine, the utterance into one or more words, each of the decoded words having a start time and a stop time associated therewith (*elements 42-50 in figure 3 and/or referring to col. 6, lines 53-67 and col. 10, lines 30-47, decoded words being cue images*); capturing a plurality of images representing body movements concurrently from the one or more words in the utterance (*camera 8 in figure 2 record images of body movement concurrently with input speech recorded by the microphone 9*); buffering the plurality of images by a predetermined delay (*element 36 in figure 2*); receiving, from the ASR engine, data including the start time and the end time of a decoded word (*elements 42-50 in figure 3 and/or referring to col. 6, lines 53-67 and col. 10, lines 30-47, decoded words being cue images*); aligning the plurality of images into one or more image segments according to the start and stop times received from the ASR engine, wherein each image segment corresponds to a decoded word in the utterance *col. 6, lines 58-67, timestamp is generated by the phone recognition. Col. 10, lines 30-47, received images of body movement is processed to include decoded words (cue images) according to the timestamp*); and presenting an image segment with a corresponding decoded word (*Display 18 in figure 2*).

10. Regarding claims 2 and 11, Braida et al. further disclose that the image player repeatedly presents one or image segments with the corresponding decoded word by looping on a time sequence of successive images corresponding to the decoded (*col. 13, line 1 to col. 14, line 67, images are playing in a sequential order*).

11. Regarding claims 3, 12, and 14, Braida et al. further disclose a delay controller operatively coupled to the visual feature extractor, the delay controller selectively controlling a delay between an image segment and a corresponding decoded word in response to a control signal (*Frame Grabber 36 in figure 2*).

12. Regarding claims 6 and 23, Braida et al. further disclose that the body movements include at least one of lip movements of the speaker, mouth movements of the speaker, hand movements of a sign interpreter of the speaker, and arm movements of the sign interpreter of the speaker (*element 16 in figure 2, when the speaker speaks into the microphone, the lip of the speaker must be moving captured by the camera*).

13. Regarding claims 17 and 19-21, Braida et al. further disclose the step of grouping the plurality of images into image segments comprises: comparing the time information relating to the captured images with the time ends for a decoded word (*col. 11, line 48 to col. 49, line 50*); and determining which of the plurality of images occur within a time interval defined by the time ends of the decoded word (*col. 11, line 48 to col. 49, line 50*); and the step of presenting the decoded word with the corresponding image



segment generated therefrom comprises repeatedly looping on a time sequence of successive images corresponding to the decoded word (*col. 13, line 1 to col. 14, line 67, images are playing in a sequential order*).

14. Regarding claim 15, Braida et al. further disclose the step of selectively controlling a manner in which an image segment is presented with a corresponding decoded word (*Cue Speech Display 18 in figure 2*).

### ***Claim Rejections - 35 USC § 103***

15. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

16. Claims 4-5 and 16 are rejected under 35 U.S.C. 103(a) as being unpatentable over Braida et al. (US 6317716) in view of Waters et al. (US Patent No. 6256046).

17. Regarding claims 4 and 16, Braider et al. further disclose a visual detector for monitoring a position of a user (*camera 8 in figure 2*), but fail to specifically a position detector being coupled to the visual detector, the position detector comparing the position of the user with a reference position and generating a control signal, the control signal being a first value when the position of the user is within the reference area and

being a second value when the position of the user is not within the reference area; and a label generator coupled to the position detector, the label generator displaying a visual indication on a display in response to the control signal from the position detector.

However, Waters et al. teach a position detector coupled to the visual detector, the position detector comparing the position of the user with a reference position and generating a control signal, the control signal being a first value when the position of the user is within the reference area and being a second value when the position of the user is not within the reference area (*col. 4, ln. 20-41*); and a label generator coupled to the position detector, the label generator displaying a visual indication on a display in response to the control signal from the position detector (*col. 5, ln. 28-59*).

Since Braider et al. and Waters et al. are analogous art because they are from the same field of endeavors, it would have been obvious to one of ordinary skill in the art at the time of invention to modify Braider et al. by incorporating the teaching of Waters et al. in order to detect the presence of users so that the system provides automated information to users in public places without human intervention.

18. Regarding claim 5, Braider et al. further disclose that the label generator receives information from the ASR system, the label generator using the information from the ASR system to operatively position the visual indication on the display (*Cue speech display 18 in figure 2*).

19. Claims 7-8 are rejected under 35 U.S.C. 103(a) as being unpatentable over Braida et al. (US 6317716) in view of Liou et al. (US 6580437).

20. Regarding claim 7, Braider et al. further discloses a display controller, the display controller selectively controlling one or more characteristics of a manner in which the image segments are displayed with the corresponding audio played (*Cue speech display 18 in figure 2*). Braida et al. fail to specifically disclose that the image segments are displayed with corresponding decoded speech text. However, Liou et al. teach that the image segments are displayed with corresponding decoded speech text (*figure 9*).

Since Braida et al. and Liou et al. are analogous art because they are from the same field of endeavors, it would have been obvious to one of ordinary skill in the art at the time of invention to modify Braida et al. by incorporating the teaching of Liou et al. in order to provide subtext for the hearing impaired individuals.

21. Regarding claim 8, Braida et al. further disclose that the display controller operatively controls the position of an image segment on the display (*Display 18 in figure 2*).

22. Claims 9 and 22 are rejected under 35 U.S.C. 103(a) as being unpatentable over Braida (US 6317716) in view of Poggio et al. (US 6250928) and further in view of Liou et al. (US 6580437).

23. Regarding claims 9 and 22, Braida et al. do not disclose that the image player displays each image segment in a separate window on a display in close proximity to the decoded speech text corresponding to the image segment. However, Poggio et al. teach that the image player displays each image segment in a separate window (*col. 10, ln. 52-67 or figure 8*).

Since Braida et al. and Poggio et al. are analogous art because they are from the same field of endeavors, it would have been obvious to one of ordinary skill in the art at the time of invention to modify Braida et al. by incorporating the teaching of Poggio et al. in order to capture viseme transitions to enable engineers study the mouth shapes created by pronouncing each particular phoneme.

The modified Braida et al. still does not disclose that each image segment is displayed in close proximity to the decoded speech text. However, Liou et al. further teach that each image segment is displayed in close proximity to the decoded speech text (*figure 9*).

Since Braida et al. and Liou et al. are analogous art because they are from the same field of endeavors, it would have been obvious to one of ordinary skill in the art at the time of invention to modify Braida et al. by incorporating the teaching of Liou et al. in order to provide subtext for the hearing impaired individuals.

### ***Conclusion***

The prior art made of record and not relied upon is considered pertinent to applicant's disclosure. Wactlar et al. (US 5835667) teach a searchable digital video

library and a method for using the library that is considered pertinent to the claim invention.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Huyen X. Vo whose telephone number is 571-272-7631. The examiner can normally be reached on M-F, 9-5:30.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Wayne Young can be reached on 571-272-7582. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

HXV

\*\*\*

7/29/2005

W. R. YOUNG  
PRIMARY EXAMINER